DOCUMENT RESUME

ED 137 686 CG 011 272

AUTHOR Boruch, Robert F.
TITLE Record Linkage in Longitudinal and Correlational Research: Its Justification and Implications for Individual Privacy.
INSTITUTION Northwestern Univ., Evanston, Ill. Dept. of Psychology.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Sep 76
CONTRACT NIE-C-74-0115
NOTE 44p.; Paper presented at the Annual Conference of the American Psychological Association (84th, Washington, D.C., September 3-7, 1976)

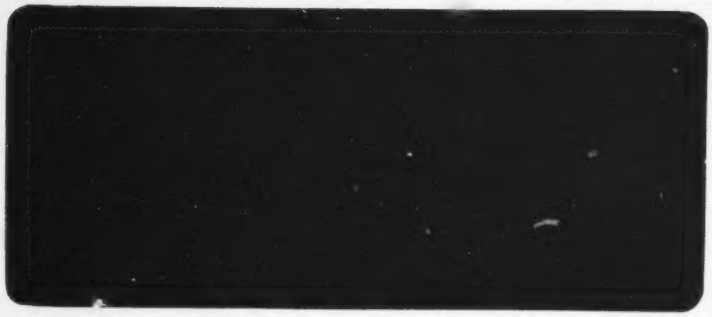
EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Confidentiality; *Correlation; Data Analysis; Evaluation Methods; *Longitudinal Studies; *Research Problems; *Social Science Research; Surveys
IDENTIFIERS *Record Linkage

ABSTRACT

This document discusses social survey research in which the need for identification of respondents may bring social research into conflict with the law and social custom. The paper deals with two features of the conflict--the products of such research, and the way in which privacy of the respondent can be assured regardless of the product. The paper is divided into five parts: (1) Introduction; (2) Longitudinal Inquire: Its Definition, Justification, and Bearing on Record Linkage; (3) Correlational Research: Definitions, Justification, and Relevance to Record Linkage; (4) Privacy Implications: Private with Respect to Whom?; and (5) Competing and Conjoint Approaches to Assuring Confidentiality of Response in Social Research. (Author/JLL)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED137686



U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EVALUATION RESEARCH PROGRAM
DEPARTMENT OF PSYCHOLOGY
NORTHWESTERN UNIVERSITY
EVANSTON, ILLINOIS

CG011272

TITLE: RECORD LINKAGE IN LONGITUDINAL AND CORRELATIONAL RESEARCH: ITS JUSTIFICATION AND IMPLICATIONS FOR INDIVIDUAL PRIVACY[1]

AUTHOR: Robert F. Boruch
Northwestern University

DATE: April, 1976

PREPARED FOR: Invited presentation to the Swedish Council for Social Science Research: Symposium on Personal Integrity and Data Protection Research. Hässelby Palace, Stockholm, March 15-17, 1976. Also, APA, September 1976.

PURSUANT TO: NIE-C-74-0115

EVALUATION RESEARCH PROGRAM
PROJECT ON SECONDARY ANALYSIS

NORTHWESTERN UNIVERSITY
EVANSTON, ILLINOIS 60201 USA

1. Introduction

In social survey research, the respondent's identification ordinarily serves as an accounting device. Both identification and associated data are maintained under the proviso that they will be used only for research purposes and, in particular, will not be used to make personal judgements about identified individuals. Despite the proviso, the need for identifiers may bring social research into sharp conflict with law and social custom. This paper deals with two features of the conflict--the products of such research and the way in which privacy of the respondent can be assured regardless of the product.

In particular, Parts 2 and 3 concern longitudinal and correlation research in which identifiers are normally deemed essential. There is a special emphasis on the practical consequences, including loss and distortion of information, engendered by thoughtless abridgement of one's ability to track individuals over time. And because the social benefits of the research will often clearly offset the privacy depreciation effects, there is a special emphasis on benefits of the research product. Our illustrations are taken from medical studies, economics, education, psychology, and sociology.

Part 4 briefly covers a topic which is already familiar to some of you--the privacy problems implied by social research efforts in general, and by the illustrations in particular. In Part 5, some general strategies for resolving problems are laid out, together with a few examples of their application. Here too the discussion is brief but broad in its coverage of procedural, statistical, and law-based solutions to the problems. The main theme here is minimizing degradation of privacy without preventing good research designed to better understand human behavior.

2. Longitudinal Inquiry: Its Definition, Justification, and Bearing on Record Linkage

Longitudinal research refers here to the process of tracking a group of individuals over time to establish how the state of that group varies and, more importantly, to establish the average relation between an individual's state at one point in time and his state at some other time. For example, one may conduct a study of adults to learn not only how health status of the group changes with age, but also to understand how the individual's health at one age is correlated with status at a later age. Obtaining an accurate characterization of this sort is necessary for describing and predicting health status. And it is crucial for the more demanding task of explaining the bio-social mechanisms which underlie health status development.

Usually, this methodology requires that an observation on a person at a particular time be linked with observations made on that person at subsequent times, for each person in a sample. The vehicle for linkage is typically, though not always, the individual's identification. The

linkage implies some degradation of privacy, and so it behooves us to ask why such research is justified, to ask what we can learn or have learned from such research.

In the following remarks, some evidence bearing on these questions is presented. Section 2.1 covers some of the logical traps in which we can easily be ensnared if we choose not to do longitudinal research. Sections 2.2 and 2.5 consider a few discrete examples of longitudinal research and their products.

2.1 Traps, Artifacts, and Circularity

One of the simplest ways to illustrate why longitudinal data may be essential for even primitive understanding is to compare it with a (ostensibly equivalent but) less demanding mode of data collection. Cross-sectional studies, for example, have been suggested as a way of learning as much about human behavior as longitudinal investigations. And because they involve observation of a large sample at only one point in time, they are said to degrade privacy to a lesser degree than the longitudinal approach.

Consider, for example, the problem of understanding how intelligence (or certain intellectual achievement) varies with age. One might conduct a survey of a sample of children of age 3, say, and then continue to survey those individuals annually until they reach an advanced age. Or, in the interest of saving time and perhaps on privacy grounds, we might choose to conduct a single survey of a representative sample of (anonymous) 3-year-olds, a sample of 4-year-olds, and so forth at only one point in time, under the assumption that this cross-sectional survey would yield roughly the same results as the longitudinal survey. This last assumption, that a growth curve based on longitudinal data will be roughly equivalent to a growth curve based on cross-sectional data, is critical.

The assumption also happens to be wrong with alarming frequency. In particular, its espousal by some human-development experts has led to some erroneous, not to say embarrassing, folklore about the development of human intelligence. The same assumption has been a trap in some economic welfare research, in some epidemiological work, and in other areas.

To understand one of the logical traps here, consider Figure 1a, a chart commonly used during the 1940's and 1950's to illustrate the gradual increase in IQ from childhood to early adulthood, and the gradual decrease thereafter. The implication of the graph, which is based on actual cross-sectional data, is that at age 30 one's IQ is at its peak, and things go downhill soon after that. What makes the chart much more persuasive is that similar inverted-U patterns show up in other investigations of human ability based on cross-sectional data. This includes the quality of treatises written by eminent philosophers (rated by eminent philosophers) when plotted against the age at which the author wrote the document. And it includes the level of innovativeness of theory and invention of chemists when plotted against the chemist's age at the theory's production, and

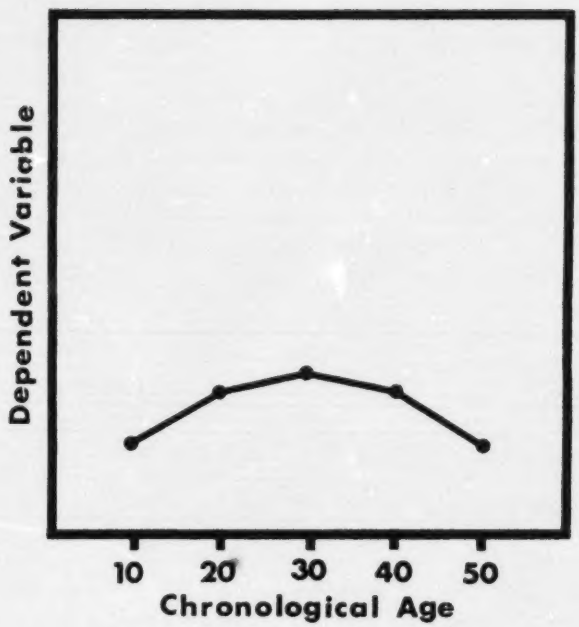


Figure 1a

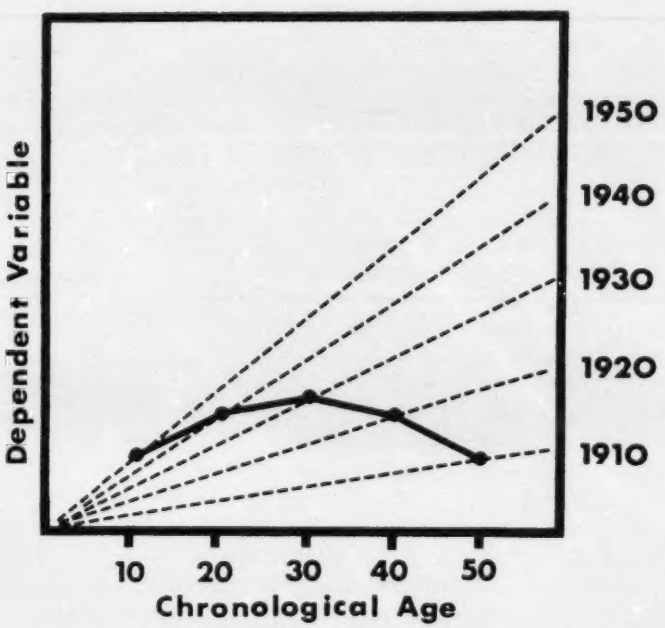


Figure 1b

Figure 1. Examples illustrating the confounding of age and cohort differences in cross-sectional research. Source: J. R. Nesselroade and Paul B. Baltes, Adolescent personality development and historical change: 1970-1972. Monographs of the Society for Research in Child Development 1974, 39, (1, Serial No. 154), page 4.

similar data (see, for example, Birren, 1964).

Suppose now that instead of the cross-sectional data, there existed longitudinal data on exactly the same individuals. The dotted lines in Figure 1b illustrate how actual IQ may increase consistently with age without a notable decline, and how the rate of increase can depend on year of birth, i.e., on cohort. The points connected by the solid line correspond exactly to what appears in the plot of cross-sectional data. The chart suggests that individuals born in 1910 increase in intelligence as they grow older. But their rate of increase is lower than the corresponding rate for a younger cohort, e.g. individuals born in 1930. The reasons for differences in development rate, or "cohort effects," are a matter of speculation. They may involve any number of bio-social factors; the differences may even be an artifact of the increasing reliability or culture-relatedness of such tests. Regardless of the reasons, the point is that the longitudinal data offer us a less misleading picture of human development than the cross-sectional data. Moreover, the theory generated by the former will differ markedly from theory generated by the latter. It is clear that relying solely on the cross-sectional data can lead one to a conclusion which is quite contrary to the way nature behaves. In point of fact, there is reliable evidence from studies by Barton et al. (1975), Schaie (1965), and others that Figure 1b is a more realistic portrayal of nature than Figure 1a.

> Exactly the same inferential problems occur for a variety of physical and social measures of individual characteristics. Plots of height, for example, when plotted against age, often show an inverted-U pattern if based on cross-sectional data, simply because rates of growth and upper limits on growth are quite high for children recently born, relative to the growth rate and upper limits for those born 80 years ago. Plots of cross-sectional data on level of extroversion and age of adolescents in certain areas of the United States make it appear that extroversion declines through adolescence when it actually increases on the average and increases most quickly for recently born cohorts. Longitudinal data on adolescent tough mindedness (autonomy, assertiveness) suggests a fair degree of stability over ages 12-15; more recently born cohorts generally exhibit higher levels of trait. But cross-sectional data show a declining trend.[2]

Some of you may regard "soft" social data, like psychological measures, as particularly susceptible to the inferential trap just described. The fact is that even data on "hard" social variables, such as income, are no less immune to the problem. Consider, for example, estimates of lifetime income for individuals. These predictions are important in the commercial arena, e.g. in some credit and loan research in the insurance business. And they are no less important in the government sector, e.g., in planning social security benefits and the like. Often there is a choice between using cross-sectional data or using longitudinal research, and if both provide equally accurate estimates, then one might choose

the cross-sectional approach for managerial reasons or on grounds that a cross-sectional survey involves less degradation of the privacy of individuals since one can presumably elicit anonymous responses. Miller and Hornseth's (1970) attempts to estimate lifetime income for certain segments of the population, is interesting in this respect.

> That estimates of lifetime income based on the two kinds of data will not be the same is clear from Tables 1 and 2. Table 1, based entirely on cross-sectional survey, suggests that annual income increases up until age 35, stabilizes during the 35-54 year age interval, then declines. The pattern is similar whether one considers data collected in 1947, or 1948, or 1949. Table 2, on the other hand, is based on longitudinal data and illustrates a much less drastic pattern, notably that increases in income persist over a wider age range, and rates of increase are substantial. The longitudinal data are, of course, affected by inflation and other factors uniquely associated with a given cohort, but similar patterns occur after adjustment for inflation. They are more accurate than the cross-sectional data in the crude sense that they better describe the way observable income behaves as a function of age.

Though the example is recent, the problem of estimating lifetime earnings from cross-sectional data is not a new one for economists. Klevmarken (1972) gives a tidy and brief description of the history of the problem in this context and points out practical needs for better estimates in labor negotiation, actuarial sciences, and elsewhere. More important, he has managed to show, using both longitudinal and cross-sectional data, how one could develop less misleading models of lifetime income curves if one had available only the cross-sectional data. He makes the same point as we do, however, in observing that there is no <u>generally</u> reliable way to establish longitudinal trends from cross-sectional data alone. Any attempt to do so must be based on assumptions which, for the social scientist, may easily fail to be met in reality.[3]

A different but no less important trap is the failure to recognize that longitudinal rather than cross-sectional data may be essential for detecting subtle influences on human behavior. The problem of designing precise investigations is particularly important in estimating the impact of social programs whose effects, we know, are often weak but may nonetheless be politically important. Achieving that objective often depends on the availability of longitudinal data. There is a large array of analytic techniques, for example, which employ the correlation between behaviors at different points in time to expunge irrelevant variation from the data. The use of longitudinal research techniques, especially in conjunction with randomized experiments, usually makes it easier to detect influences which might otherwise be obscured by the normally high variation in human behavior.

> Consider, for example, the Cali, Colombia experiments on the impact of nutritional supplements on children's physical

Table 1

Estimates of Mean Annual Income in Dollars for Age 25 through 64, Based on a Cross-Section of Men Sampled in 1947, a Cross-Section Sampled in 1948, and a Cross-Section Sampled in 1949

Year/Age	25-34	35-44	45-54	55-64
1947	2704	3344	3329	2795
1948	2898	3508	3378	2946
1949	2842	3281	3331	2777

Adapted from data presented by Miller and Hornseth (1970).

Table 2

Estimates of Mean Income in Dollars over 10-Year Intervals for Six Cohorts of Individuals

Year/Age	25-34	35-44	45-54
1. 1947	2704 (1947)	5300 (1957)	8342 (1967)
2. 1948	2898 (1948)	5433 (1958)	8967 (1968)
3. 1949	2842 (1949)	5926 (1959)	9873 (1969)
Year/Age	**35-44**	**45-54**	**55-64**
4. 1947	3344 (1947)	5227 (1957)	7004 (1967)
5. 1948	3508 (1948)	5345 (1958)	7828 (1968)
6. 1949	3281 (1949)	5587 (1959)	8405 (1969)

Note. Each cohort has been surveyed every 10 years. The first cohort, for example, contains individuals who were 25.34 years of age in 1947 and had an average of $2,704; in 1967, when they were 45-54 years of age, their mean income was $8,342. Adapted from Miller and Hornseth (1970).

> growth. Special nutritional supplements were assigned randomly to a sample of malnourished children; supplements, which were in short supply, were unavailable to an otherwise equivalent sample of comparison group children. The impact of the supplements was not evident from scrutiny of mean changes in treated and untreated groups; the simple natural variation in heights of even malnourished children is sufficiently large to obscure real differences. More sophisticated analyses, using correlations between repeated measures of height of the children, did yield estimates of program effect which differed notably from chance level. As a consequence of the positive finding, the supplements are being improved, put into local production, and tested on a much larger scale in three other less well developed countries. (Bejar, 1975; Sinesterra, McKay, & May, 1971)

The same use of a longitudinal approach for the sake of sensitive analysis of program effects is evident in other areas. Heber et al. (1972), for example, have conducted 6-year studies to determine the relative impact of special programs for reducing the risk of functional retardation among infants and young children; based on these Wisconsin pilot tests, similar test programs are being mounted in North Carolina and elsewhere. Beyond the midpoint in Kaiser Permanente's 10-year experiments, Ramcharan et al. (1973) find evidence for the impact of multiphasic screening on prevention of disease, an impact which is bound to be negligible during the first few years of the program. In the economic area, the Housing Allowance Experiments require 8-12-year followups to determine incremental benefits of income subsidy plans on the poor, and to provide information for effective legislation in the area. In these cases and in innumerable others (see Riecken et al., 1974), the effects may be undetectable in the short run, and difficult to detect in the long run, especially if the groups involved are quite small. There is simply no reliable substitute for longitudinal followups in these instances.

The final logical trap of interest here bears on both longitudinal and correlational research; it involves the analysis of data based on aggregate of individuals in order to make judgements about individuals within the groups. To establish the average relation between literacy and race in the United States, for example, one might obtain published census statistics on the proportion of literate persons and the percentage of Negroes for each of 48 states and then compute the correlation between the two variables. Aggregated data might be used here on grounds that the relevant information is easily accessible from published tables. Or, we might justify our action on grounds that the use of published data does not present the privacy-related problems which might be engendered by a special survey.

There are two weaknesses implicit in the argument that aggregate data can be used in lieu of individual data. The more obvious one is that inferences made about groups are not necessarily appropriate to the individual and in fact may be quite inaccurate. The second weakness, more

a matter of precision than accuracy, is that analyses based on grouped data are often considerably less likely to reflect changes in individuals than analyses based on data at the individual level.

To be more specific, consider the literacy-race example. At a particular point in time the correlation between literacy rate (percent literate) and color (percent Negro) computed on the basis of the nine census regions of the United States is .95. When individuals are grouped by state rather than region, the correlation is .77. Finally, when individuals are not grouped at all, but the entire disaggregated population is considered, the correlation is .20. (The example is from Robinson's fine paper (1950) on census data prior to 1950.) A similar problem with a different resolution appears if we try to determine the relation between color (white-nonwhite) and occupation (domestic service--other) for female employees in Chicago in 1940. Though a correlation based on percentage data for each of nine areas is .34, the actual correlation based on individuals is .29, not too different from the area-based estimate (see Duncan & Davis, 1953; Goodman, 1953).

In the literacy-race example, the high correlation obtained from the regional data might be interpreted as suggesting that illiteracy is pervasive among blacks, and furthermore, that a massive program of education must be put into effect to counteract the problem. In point of fact, if we look at individuals' data, rather than at data based on opportunistic groups into which individuals may fall, we reach a considerably less pessimistic and a more accurate conclusion: that the relation between race and literacy was small but notable. Any attempt to resolve the problem of illiteracy by making a massive investment in rehabilitating the reading skills of each individual based on the .95 regional correlation, is bound to be a wasteful allocation of scarce resources.

An obvious problem in these matters is the use of aggregates of individuals as a surrogate for individual persons. Since the aggregates are usually constructed for political or administrative purposes (e.g. census regions, health care service regions), it is unlikely that these "natural" aggregates will constitute valid replicas of real persons. We have only a little theory to guide us in selection of "proper" aggregates. And it is impossible to predict whether an aggregate will be proper without some data at the individual level.

The problem is a chronic one in the social and administrative sciences which must rely heavily on aggregated data--sociology, epidemiology, economics, statistical geography. It is particularly crucial in attempts to evaluate the impact on national social programs on individuals. Many such evaluations, in education for example, rely on data aggregated at the school district level to estimate the impact of a nationally supported compensatory reading program for disadvantaged youth. The inferences made about individuals (based on analysis of aggregates rather than on individuals) are generally biased in an unknown fashion (the individual data not having been analyzed), and are imprecise because the aggregate data are insensitive to changes, even some marked

changes, in individuals (see Burstein, 1976, for examples). This is not to say that the aggregation problem will always yield biased estimates. It is to say that the problems are crucial and cannot be resolved unequivocally without some evidence based on individual rather than aggregated data.

2.2 Medical Research

There is a fine tradition of longitudinal studies in medical research, dating at least from Hippocrates's efforts to characterize the progressive stages of disease among his own and his colleagues' patients (King, 1971). The systematic tracking of both the healthy and the ill remains a basic weapon in medical research armamentarium. At its best, the approach not only helps to identify the existence and incidence of a disease entity, to determine symptom development and disease consequence, but it is essential in laying out the array of possible progenitors of the disease. Longitudinal methods in this sector have become considerably more efficient over the last 40 years with the development of survey sampling technology. And when coupled to other methods, such as randomized experimental tests, the approach can be dramatic in identifying whether and how well particular treatment programs work.

Examples of the process are not hard to find. But for the sake of detail, suppose we examine a complex research area which, by virtue of gifted science writers (such as Gilmore, 1973) and researchers (such as Kannel et al., 1961), is among the best documented. Modern work on coronary heart disease appears to have reached a turning point during the 1940s and 1950s with autopsy studies. Those investigations, because of their small size and cross-sectional nature, provided thin support for the linkage among natural development of arteriosclerosis, heart disease, and bio-physical conditions (blood pressure, etc.), and more importantly, provided the evidence necessary to justify longer term longitudinal study of the problem. The Framingham Study (Kannel et al., 1961), among the largest of subsequent efforts, was designed to better establish relations between prior condition and subsequent death due to heart attack. Spanning 25 years in the lives of 9000 men, the effort was of sufficient size and duration to permit computation of risk factors operating in the population: Actuarial tables were developed to illustrate the likelihood of heart attack as a function of earlier serum cholesterol level, blood pressure, EKG abnormalities, and so forth. Other studies--animal experiments and comparative investigations of populations with natural differences in these factors--yielded evidence which added to speculation about the role of serum cholesterol level and other factors in heart disease. Because the ability to describe and predict based on longitudinal does not necessarily yield unequivocal information on causes of the disease, and because study of human population yields results which are similarly ambiguous, long-term, experimental tests of alternative treatment programs have been mounted. The best of those tests generally involve large samples tracked over long time periods and, moreover, randomized assignment of individuals to one of the competing treatments. As a consequence,

they raise problems more serious than those engendered by longitudinal research alone. Nonetheless, pilot efforts, such as the Diet Heart Feasibility Study, have been completed to furnish data on the practical difficulty of field tests and somewhat less equivocal small-scale data on the impact of diet control on heart disease. Such short-term (two-year) studies have paved the way for longer term studies which focus on the more plausible and tractable causal mechanisms, notably reduction of heart disease through diet or drugs which reduce serum-cholesterol levels. The largest of current clinical trials will run five years and involves over 50 institutions and 8000 patients; it is designed to evaluate the effectiveness of alternative drugs and drug dosage level for reducing cholesterol level in the bloodstream (Coronary Drug Project Research Group, 1973). Although the primary response variables are mortality rate due to heart disease and related illness, a variety of social, biological, and physiological measures are being obtained. The social measures--smoking habits, lifestyle measures, race, job characteristics, and so on --are expected to add precision to results and to help identify variables which though influential are less amenable to direct control.

The products of earlier longitudinal studies coupled to experimental tests are readily accessible (see Boruch & Riecken, 1975; Riecken et al., 1974, and references cited therein). Long-term followup of released prisoners who have had cosmetic surgery to remedy facial disfigurement has given us evidence for lower recidivism rates among prisoners so treated. Longitudinal experiments on the effectiveness of physician surrogates--nurse practitioners, physician extenders--has yielded information essential for reducing costs of medical service, for planning innovative programs in health care utilization, and the like. A new generation of pharmacy research focuses on both short-term and long-term drug-taking behavior--determining level of patient compliance with medication regimens, determining how special packaging of medication influences compliance especially among the elderly, and so on. In preventive medicine, tests of the impact of multiphasic screening by Kaiser-Permanente, are being run for a 10-year period to assure that long-term effects of annual screening on detection and amelioration of disease are well documented.

In brief, these examples and others like them teach us that there is no way to establish etiology of disease or to evaluate the effectiveness of prevention and treatment programs without longitudinal study. Not that longitudinal research is sufficient. Its natural limitations must usually be broadened by coupling this approach to others, notably experiments, designed to establish cause-effect relations. But the idea is central to medical research and, in principle and in practice, generalizable to other areas.

2.3 Psychology and Psychiatry: Biochemical Bases of Schizophrenia

For the past 100 years, the scientific and lay arguments over the causes of schizophrenia have been supported largely by ambiguous data. The information at its worst has been unreliable and no more than anecdotal in form; at its best it has been based on longitudinal study of very small

numbers of individuals and heavily reliant on retrospective reports of unknowable reliability. The debate's focus has changed markedly during the years, however, in part because of longitudinal research which depends heavily on record linkage (Mednick & McNeil, 1968; Mednick, Schulsinger, & Garfinkel, 1975).

One of the basic problems in discovering the origins of schizophrenia, as many of you know, is to disentangle the biochemical causes of the problem from the environmental influences. To resolve the problem, researchers at Denmark's New School for Social Research, at the Psykologisk Institute (Copenhagen), and at the Kommune Hospitalet (also Copenhagen) have conducted longitudinal studies of over 4000 adopted children to discover how incidence of schizophrenia among them varies with occurrence of schizophrenia in their natural families and in their adopted families. If, for example, the schizophrenia among children born of schizophrenic parents but raised by adopted nonschizophrenic parents is high, then one has more reason to believe that the malady's origin has a genetic component. Schulsinger's findings, obtained in collaboration with David Rosenthal, Seymour Kety, and Paul Wender of the United States are that:

- The incidence of schizophrenia is substantially higher among adopted children who had schizophrenic parents than among adopted children whose foster parents were schizophrenic.

- There is a very low incidence of schizophrenia among children born of parents who are not schizophrenic.

- If children whose natural parents are not schizophrenic are later adopted by a schizophrenic foster parent, there is no increase in linkelihood that the child will become schizophrenic.

This information is an elementary but important step in establishing the credibility of the idea that the origins of schizophrenia are partly environmental and partly genetic, and it is important in directing attention to more fertile areas of research. The latter include careful studies of the possible genetic mechanisms and of the role played by certain enzymes (for example) which may produce a predisposition toward schizophrenic behavior.

These findings could not have been made without longitudinal data on adopted children and their natural and adopted parents, and without the crucial linkage among existing medical records, social service records, and follow-up data collected more recently on the basis of the national address registry.

2.4 Longitudinal Study in Manpower Economics

In human resources research, good evidence for the usefulness of longitudinal data has been scanty in part because the relevant data have

been in short supply. The recent buildup of longitudinal files has helped greatly to understand the data's benefits and limitations, however. Of particular interest are the National Longitudinal Surveys (NLS) of the U.S. labor market, begun in 1966 by Herbert S. Parnes (1975). Those data are based on repeated surveys of a national probability sample of 20,000 individuals in four labor market strata: middle-aged men (45-59 years old at the survey's beginning), women (30-44 in 1966), young men and young women (14-24 in 1966). The resultant data are being updated periodically and, stripped of identifiers, are being made available to the community of manpower researchers. Aside from their obvious benefits for temporal description of the labor market, the data can be very informative on account of their longitudinal feature. Parnes (1975) maintains that:

> Perhaps the single most important contribution of longitudinal data is that they facilitate the identification of causal relationships that cannot confidently be identified in any other way. Take, for example, the relationship between attitudes and behavior. In cross-sectional data, such relationships are ambiguous, since one cannot be certain whether the attitude produces or reflects the behavior. Does job dissatisfaction lead to turnover, or does an association between the variables simply mean that individuals who quit jobs are likely to rationalize their behavior by reporting (retrospectively) that they were unhappy? When attitudes measured at one point in time can be related to <u>subsequent</u> behavior, such ambiguity disappears. The NLS data for middle-aged men have clearly demonstrated that the degree of job satisfaction <u>predicts</u> the likelihood of a voluntary job separation and that a commitment to work in general, as well as satisfaction with one's particular job, decreases the likelihood of early retirement.
>
> The usefulness of longitudinal data in clarifying causal relationships is, of course, not confined to instances in which one of the variables is attitudinal. For example, finding that the receipt of training by middle-aged men between 1966 and 1971 was associated with a net earnings advantage in 1971 (controlling for such other factors as education, health, and region of residence) Avril Adams went on to demonstrate that the trainees-to-be had already enjoyed higher earnings in 1966 (again controlling for the same variables). Thus training was found to be a selective process, presumably attracting the more highly motivated or otherwise more productive individuals. To put the matter differently, some part of what would doubtless have been identified by a cross-sectional analysis as training's contribution to earnings was found to have reflected an incompletely specified model--i.e., the failure to control adequately for factors associated both with earning and the probability of receiving training. (Parnes, 1975, pp. 246-247)

Professor Parnes is optimistic about the fruits of research based on his data files. We do not share that optimism, since longitudinal data alone is often insufficient to make unequivocal judgements about the impact of manpower training programs. We do agree that such data are essential for better understanding and prediction of gross labor market behavior, for establishing tentative hypotheses which can be later verified using more controlled studies, and for prediction.

2.5 Longitudinal Study in Child Measurement

The National Child Development Study (NCDS) began in 1958 with a survey of some 5,000 pregnant women. Its main objective (like the United Kingdom's earlier study, the 1946 Population Investigation Committee survey) is to establish the linkages among prenatal conditions, environmental factors, and growth of young children. According to Wall, the results of 1966 followup data on over 9300 children showed, despite strong suspicion to the contrary, the following variables are not singly predictive of lowered reading ability: maternal hypertension, breech presentation or forceps delivery, Caesarean section. Further, there is an unexpected and strong relation between departure from normal gestation and reading and social adjustment test scores at age 7; gestational maturity is a better predictor than the more commonly accepted birth weight measurement.

The Population Investigation Committee Study yielded other conclusions which could not have been reached without longitudinal data. Taken verbatim from Wall and Williams (1970, pp. 42-43), we have:

> That the effects of social mobility and increasing material prosperity have differential effects according to the educational levels of the parents and the number of children in the family.
>
> That a relatively poor social environment is cumulative in its effect on children's height, girls being more sensitive to this than boys.
>
> That separation from mother, as well as being much more prevalent than had been thought, seems (in the period from birth to five years) to provoke less serious permanent disturbance than might have been expected from clinical studies of a post hoc kind.
>
> That by the age of 5, broken homes apparently do not provoke more than temporary disturbance (bedwetting), and this only in non-manual families.
>
> That the proportions of mothers taking up full or part-time work increased as their children approached the age of 5, but that there was no evidence that their children were less emotionally stable at this age.

That early toilet training leads to earlier bowel control, less bedwetting, and less breakdown later.

That a high proportion of bedwetters bite their nails and have speech defects, difficulties which persist even after they become dry.

That children prematurely born are more vulnerable physically during their first two years but not afterwards; that although rather smaller than normal children in later childhood, this is not the result of prematurity, and that by the age of 8 they tend to be handicapped in mental ability, particularly in reading. (Wall & Williams, 1970, pp. 42-43)

2.6 Education and Its Impact

For the sake of better allocation of scarce resources to education, it is reasonable to learn how education affects academic achievement, and subsequently, earnings. We need to know what the most effective elements of the education process are, how they work, and how they affect the individual's intellectual and economic development. The impact question is especially relevant to novel programs designed to overcome disadvantages under which some social groups labor, i.e., designed to introduce more equity into the social system through education.

Most research designed to get at these issues begins with cross-sectional surveys and even a brief incursion into history shows that these have been useful despite the limitations of the cross-sectional approach. Abraham Flexner's 1910 report of his studies of medical schools in the United States relied solely on this methodology and on Flexner's standards of performance to produce a major reformation in medical training. The Thorndike and Ayres studies of school record systems were similarly useful in moving schools toward higher quality (though still imperfect) record-keeping practices (Goslin & Bordier, 1969) The cross-sectional studies have been, and still are, enormously useful in this context, especially in clarifying the scope of educational problems, especially where standards of quality are fairly clear.

But the difficulty of making inferences about the impact of education, of understanding individual growth, based on cross-sectional data, are no less severe in this sector than they are in the medical arena. It is difficult, often impossible, to discriminate accurately between the influence of background variables and those of the school. It is not generally possible to lay out growth and assay impact without at least some longitudinal data. The scientific and political traps here are exemplified by the current U.S. controversy over busing students from the school district in which they live to another in the interest of fostering equitable, quality education. James Coleman's advocacy position five years ago, based largely on cross-sectional data, is considerably different from his opposition now based on longitudinal data and demonstration projects.

Better interpretation, inference, and prediction are conditional on better theories (models) for simulating social behavior and on data necessary to support that theory. As a consequence of the shortcomings of earlier data, a number of longitudinal studies have been mounted to better understand education impact. We cannot summarize those here--there are far too many to do so reasonably. So we content ourselves with examining a nice study by Fägerlind.

This analysis of longitudinal data was designed, in part, to clarify polar views of the results of public investment in education: that duration of educational experience has a major import on earnings, a view exemplified by Nobel Laureate Paul Samuelson, and Jenck's (and others') view that the impact on earnings of education beyond the post-secondary level is marginal. The Fägerlind research manages to avoid the traps of cross-sectional data and of short-term longitudinal study, by considering individual growth over a 30-year-period. It is an efficient study of a well defined social group insofar as it builds on survey data initially collected in 1930 on a subpopulation of children in Malmö. It obtains both economy and completeness of sampling through the use of population registries for follow-up surveys. Accuracy and temporal relevance of data are enhanced by relying on archival records--military selection test-scores for men, data from tax registries on earnings of the respondent and the respondent's parents, and data from census records on demography, geographic and occupational mobility, etc. Fägerlind supplemented archival records with survey data collected during the 1940's, '50s, '60s, and early '70s.

The product of this particular research is interesting not only in adjudicating polar views: Fä gerlind's data, of higher quality than Jencks', supports Samuelson's theory. It also helps to specify the process, the mechanism underlying education's impact on earnings through an otherwise tangled mass of competing influences such as home, family, and so on. And it has helped to understand shortcomings of competing data and models: quality of education, for example, has been ignored in many such analyses and this one uncovers strong, plausible linkages between quality and earnings from age 30 onwards.

Still, the longitudinal approach used here is only an int..im step. It is naturally limited in the extent to which it can be applied in specific, particularly novel settings. More recent research, for example, stresses small longitudinal experiments, mounted alone or in conjunction with larger observational studies, to obtain finer appraisals of innovative educational programs and practices. Some, like the Heber et al. (1972) work is dedicated toward inhibiting intellectual deprivation from infancy. Others, like Middlestart, involve randomized tests of programs designed to improve academic performance of adolescents who are unusually deprived by virtue of their very poor economic condition. Still other experiments, designed to improve medical school education, police training, and manpower training and the like, follow participants through adulthood in the interest of obtaining less ambiguous information about the short-term impact of expensive and specialized education programs (see Boruch & Riecken, 1975; and Riecken et al., 1974).

3. Correlational Research: Definitions, Justification, and Relevance to Record Linkage

Correlational research refers here to the process of establishing how two characteristics of an individual are related to one another. The average relation, for a large sample of individuals, may be represented in statistical form by a simple correlation coefficient, by the probabilities in an actuarial table, and so on. For example, to identify the relation between level of health status and level of physical activity during work, one might obtain measures of both variables from each member of a suitable sample of individuals, link the two elements of information on each individual, then compute an index of the relation based on that linkage. The correlation may be of descriptive interest alone in that it reflects the existence and strength of a relation between two variables. It may be more important to an individual, in that the correlation helps to predict future health status from current physical exertion levels. Finally, such data make it possible to form tentative ideas about the biochemical mechanism by which exertion influences health status (or vice bersa), i.e., to build theory necessary for the development of better control of health status.

In principle, correlational investigation is a general activity of which longitudinal research is an important subclass. Both types of research usually require some form of record linkage to sustain statistical analysis. They are discussed separately here on account of traditional differences in the emphasis of each type of research.

Correlational research often requires that the contents of records which are maintained by independent archives be linked. The special functions of linkage vary considerably, but most can be grouped into one of the following categories:

- To assess and improve the quality of available data from any source;
- To reduce costs, duplication of effort, and respondent burden in surveys;
- To clarify and enrich the data base for applied social research and policy analysis.

The illustrations of the benefits and limitations of linkage are presented below using this taxonomy to organize our experience.

3.1 Assessing the Quality of Data and Improving the Quality of Data Analysis

"Response validity" refers to the association between an individual's response to inquiry under one set of conditions and his response to inquiry under a second set of conditions which are thought to facilitate near-perfect reporting. Most such studies involve one form or another of

record linkage. Census data on income of identifiable respondents may be linked to Internal Revenue Service reports, for example, to assay the adequacy of the census interview process. Data from interviews made on one occasion under normal conditions may be linked similarly to later, more intensive, interviews to gauge the adequacy of the "normal" interview conditions. Some mechanism for linkage is critical for computing quantitative indices of average agreement between the two types of reports.

Without some empirical basis for judging the data's credibility, it is impossible to lend any meaning to statistical analysis, unless of course, one is willing to cover the whole matter with a secular act of faith. The absence of validity statistics is especially crucial not only in interpreting descriptive statistics but also in using them to monitor and evaluate social programs. Errors in reporting will usually make it more difficult to detect changes in human status, and in situations where data imperfections go unrecognized, data analysis may result in wildly inaccurate conclusions.

Examples from descriptive survey research. Many of the better validity studies in the U.S. have been conducted by government agencies and by university-based research groups. The studies are frequently designed to furnish sufficient evidence to support an administrative decision about whether or not to continue a particular type of inquiry.

In health survey research, for example, a good deal of the use of record linkage is reported in the proceedings of a recent national conference (Reeder et al., 1975). The deficiencies in physicians' records, for example, have been examined by matching record content with data from interviews with patients. Distortions in reports made by physicians to their own medical societies have been investigated by linking those reports with intensive interviews subsequently conducted with physicians themselves. Methods of interview designed to minimize embarrassment in health-related surveys have been tested and evaluated using individuals' hospital records as the standard for accuracy. Surveys of health services utilization, necessary for planning such services at the national level have been validated using side studies which link individual responses to records maintained by providers and third-party payers.

Analogous examples appear in manpower research. For example, to appraise the validity of self-reported "occupation five years ago," a question which has appeared in many cross-sectional manpower surveys, the U.S. Census Bureau conducted tests on 2800 households in 1968, for whom 1963 data on actual occupation were available. Despite the use of a variety of methods to elicit the retrospective report, the differences between retrospective report and actual status were in the range 23-28% (Jabine & Rothwell, 1970). The linkage here,

between 1963 archival records and the 1968 survey, was essential in establishing the validity rate and in the pattern of invalidity. And the statistics themselves influenced the Census Bureau's decision to drastically reduce the use of the retrospective question in its own surveys, and to routinize the correction of other survey researchers' occupational mobility statistics.

Housing statistics are no less immune from biasing influences and, in some cases, intensive reinterviews are necessary to establish validity of initial interviews. For example, it is not unreasonable to expect that interviewers will vary notably in their ability to rate quality of housing. In testing alternative methods for assuring accuracy of the rating, the U.S. Department of Housing and Urban Development and the Census Bureau found, using reinterviews as a standard, that no particular method of interview classification yielded ratings at a reasonable validity level. And as a consequence, the rating scheme was dropped entirely in the 1960 census. Instead, crude indicators of quality (cooking facility, indoor toilet, etc.) were included in the enumerator's protocol. Again, neither the collection of validity statistics nor the subsequent administrative actions would have been possible without some mechanism for linking initial enumerator reports with more expert reinterviews.

In estimating undercounts in the census of 1960, Marks and Waksberg (1966) report both positive benefits and negligible benefits in using archival records. The use of 1950 Census records, hospital records of birth during 1950-60, records from intermediate census research, and records from the U.S. Immigration Service for special subsamples yield useful and credible evidence for under-enumeration of 2.6 to 4.7 % in the 1960 census. Similarly, for special subgroups, undercount estimates were obtained. Lists of college students were obtained from colleges to estimate undercounts in the enumerated count of 2.5 to 2.7%; Social Security addresses were used in estimating a 5.1 to 5.7 undercount in beneficiaries in the 1960 census. On the other hand, matching of census rolls against lists of relatively inaccessible individuals--lists of welfare recipients, postal service listings--"provide no special encouragement for use of matching special lists as a coverage improvement program." Horwitz (1966) conducted similar studies in rural areas which suggested that 20 to 25% under-reports in death rates and 15 to 20% under-reports of birth rates are not unusual when hospital and state medical records are used and a standard.

These examples illustrate how validity statistics, generated through record linkage, can help to delimit the credibility of social survey

statistics and can serve as a basis for making decisions about the conduct of a survey effort.

The practice of conducting side studies such as these, based on limited record linkage, is practically nonexistent in commercial survey efforts. It is, however, typical in some governmental surveys and in research conducted by some university-based research groups. That the practice is increasing even in these sectors is evident from the bibliographies published on the topic (notably Scheuren & Colvey, 1975), from new reporting systems such as Studies from Interagency Data Linkage for describing the products of the work, and other evidence.

Examples from program evaluations. Imperfections in either social survey data or administrative records make it difficult to detect and, in the worst cases, can produce statistical artifacts which make programs appear harmful. Estimates of validity, whether based on record linkage or not, are often essential for refining the design of an evaluation to accommodate the problem.

More specifically, one of the chronic problems encountered in the United States has been the production of biased estimates of program effects under some special but common conditions. Conventional statistical techniques, such as regression analysis, covariance analysis, and matching, when applied to fallible data obtained in some observational evaluations, yield consistently biased estimates of program effects, in part because imperfect measurement goes unrecognized. Consider, for example, the Westinghouse-Ohio evaluations of "Headstart," a preschool program for the economically deprived. The initial evaluation relied on a textbook application of covariance analysis of survey data to explain how children's verbal ability varies as a function of demographic characteristics of the children and of their families, and other variables. The estimates of the impact of Headstart were actually negative, implying that the program had a harmful effect. It is clear from secondary analysis of the same data that if one adjusts the conventional analysis so as to recognize imperfect measurement, the program's effect is negligible and perhaps even slightly positive (Magidson, Campbell, & Barnow, 1976). Similar biases have been discovered in the evaluation of manpower training programs (Direction, 1974), in the estimation of the imapct of special medical treatment regimens (James, 1973), and elsewhere (Campbell & Boruch, 1975).

To summarize, we observe that measures of social, psychological, medical, or economic behavior are usually imperfect. If the imperfections go unrecognized, then statistical analysis of the impact of programs designed to ameliorate relevant problems will be insensitive at best, misleading at worst. Statistics bearing on validity and reliability of response are necessary for rational adjustment of conventional statistical analyses so as to reduce bias in estimates of program impact. Record linkage is often, though not always, necessary for production of the necessary information on validity of the observations.

The view that administrative records ought to serve as the standard against which survey records are judged is, at times, clearly unjustified. Administrative records are tied to administrative action, and for that reason, are normally susceptible to a variety of biases and sources of error which do not affect survey data. One of several ways to appraise the credibility of statistics based on those records is through specialized designed surveys.

Prior to 1910, for example, studies by the noted educational researcher E. L. Thorndike on the adequacy of school records led to major reforms in school record-keeping practices. Those studies relied partly on record linkage to furnish evidence concerning deficiencies in existing record systems (Goslin & Bordier, 1969). Later studies, conducted by economists, contributed to what we now know about needs for record accuracy, publicity, and adequacy in preventing abuse of power by public utilities (see Shils, 1938). More recently, Campbell (1975) and others have tried to enumerate more fully the reasons for corruption of administrative records and to develop some crude theory to account for the phenomena. Most of the theory building depends in one way or another on the conduct of surveys to appraise the quality of an archive's contents. The U.S. Army reporting system for drug abuses, for example, were assessed during the early 1970s using an experimental interview method which generally yields less distorted information on actual abuse by identified individuals (see Section 5). The debatable quality of criminal records maintained by police has led to Federally funded victimization surveys, conducted by the Census Bureau to determine the nature and incidence of unreported crime, the elasticity in police definitions of crime, and so on. These more recent examples do not depend on record linkage to make their point. But whether a social scientific survey can be mounted to verify the quality of an archival record system depends heavily on administrative endorsement of the idea that multiple indicators of a period that are desirable. As the practice of conducting this kind of study increases, the need for more depth of inquiry and, consequently, linkages between archival record and survey record will undoubtedly increase. It is often possible to eliminate confidentiality-related problems in this context by using the insulated data bank strategy described in Section 5 below.[4]

3.2 Reducing Costs, Duplication of Effort, and Respondent Burden

Partial duplication of a data collection effort by several agencies may be justified on several grounds. Independent archives which maintain some overlapping information, for example, may be warranted by legislation which requires independent collection and maintenance of the data, they may be justified as a device for periodic cross-validation of the contents of files. Nonetheless, exact or nearly exact duplication may be costly to the data collection agencies and to the respondent who must contribute the time required to supply the information to each agency.

Although existing archival records have not often been used as a basis for evaluating the impact of experimental social programs, they do

have some promise in this regard. The argument that archival records can be used to mount more economical and more informative evaluations of social programs has been advanced persuasively by the Committee on Federal Program Evaluation of the National Academy of Sciences. We quote verbatim from that report:

> Once the major administrative archives of government, insurance companies, hospitals, etc., are organized and staffed for such research, the amount of interpretable outcome data on ameliorative programs can be increased tenfold. For example, Fisher (1972) reports on the use of income tax data in a followup on the effectiveness of manpower training programs. While these data are not perfect or complete for the evaluation of such a training program, they are highly relevant. Claims on unemployment compensation and welfare payments would also be relevant. Cost is an important advantage. Using a different approach, Heller (1972) reports retrieval costs of $1 per person for a study of several thousand trainees. Even if $10 were more realistic, these costs are to be compared with costs of $100 or more per interview in individual followup interviews with ex-trainees. Rate of retrieval is another potential advantage. Followup interviews in urban manpower training programs have failed to locate as many as 50% of the population, and 30% loss rates would be common. Differential loss rates for experimental and control groups are also common, with the control groups less motivated to continue. In the New Jersey Negative Income Tax Experiment, over three years, 25.3% of the controls were lost, compared with a loss of only 6.5% of those in the most remunerative experimental condition. While retrieval rates overall might be no higher for withholding tax records, the differential bias in cooperation would probably be avoided, and the absence of data could be interpreted, with caution, as the absence of such earnings. (Campbell et al., 1975)

It takes little imagination to see how relying on existing archival data can reduce the expense of a program evaluation. It is quite another matter to employ such records creatively in difficult research settings. One of the more clever applications of archival data stems from an effort by Robertson and others (1972) to evaluate the impact of TV messages which encourage drivers to wear their seat belts:

> In some recent tests, four different types of TV messages were broadcast over four different TV cables, each cable serving a random set of households within a large region. The research objective was to determine which TV or broadcast fostered the highest rate of seat belt usage. To evaluate usage, the researchers first observed whether or not drivers in the region wore seat belts as they stopped for lights at randomly selected intersections. To link actual usage with area of

residence, i.e., with TV message type, some mechanism for identifying each driver's residence was necessary. Rather than question each driver, the researchers merely recorded auto license numbers and employed State Motor Vehicle archives to identify the driver's area of residence. Once each driver's residence and seat belt use were linked, it was an easy matter to compare the crude effects of alternative TV messages on use.

Some examples of the savings engendered by temporary and limited linkage of governmental records have been documented by Hansen and Hargis (1966). In these cases, a sample of records maintained independently by the U.S. Census Bureau, by the Internal Revenue Service, and by the Social Security Administration were linked to determine how costs of surveys might be reduced.

Prior to 1954, for example, the Economic Census of manufacturing, retail, and other industries was conducted by field interview survey with some larger firms canvassed by mail. In the interest of reducing costs markedly, mail survey was considered as an alternative to expensive field interview surveys. At that time, the Census had no mechanism for construction and maintenance of up-to-date mailing lists, however. Such mailing lists were maintained by Internal Revenue Service and Social Security Files, based on payroll tax records, and with some modification, the basic lists were checked for validity, then adopted by the Census Bureau as a basis for the mail survey in the economic census. To obtain data on the retail industry, conventional Internal Revenue Service forms were modified slightly, making it possible to eliminate any additional mail or interview surveys of this industry by the Census Bureau. More than $6 million were saved by employing this last strategy.

Similar savings were said to have been realized in the 1967 Economic Census where, for example, modifications to Internal Revenue Service schedules permitted use of these forms to elicit necessary information, and small direct interview samples were adjoined to this effort to obtain necessary data on products, merchandise lines, and so forth. Finally, "administrative records from the Social Security Administration and from the Census) have been used to construct mailing and sampling lists economically for Bureau data collection programs and to avoid duplicating the collection of information.

3.3 Clarifying and Enriching Statistical Data for Policy Analysis and Applied Social Research

By clarification here we mean obtaining a better understanding of the meaning, nature, and limitations of a particular social statistic. "Employment rate," for example, is a deceptively simple label for a characteristic which is complex in origin. Clarification often implies an additional objective, that of enriching the data resource with respect

to number and kind of data archived, for the sake of higher quality analysis. Improving the interpretability and analyzability of a data set can be accomplished in a variety of ways. Linking of multiple data sources for statistical purposes is one method of doing so. Note, however, that linkage of all individual records may not be essential; linking a (random) sample of records is often sufficient for this purpose.

To be concrete, consider that in the United States, the Internal Revenue Service, the Social Security Administration, and the Census Bureau each independently collect data on annual income from citizens. The separation of effort is related to differences in the various agency functions. Two of the Social Security Administration's primary missions, for example, are understanding income redistribution at present and estimating the impact of redistribution policy in the future. Most U.S. citizens are required to pay a social security tax based in part on gross income, but Federal employees often do not choose to enroll in the national Social Security plan and so their incomes are not on file in SSA record systems. The Internal Revenue Service directs its attention at a different but overlapping universe, the tax-paying public, it has a different function, taxation, and it defines income differently, notably in terms of "taxable income." The U.S. Census Bureau's definition of income differs from each of the other agencies' definitions because its function is unique--statistical description of the state of the population--and because there are severe limitations on the way in which census data can be collected.

The result of these differences in definition of income, universe, and in function is that the relationships among these various sources of data on "income" have not been well understood. The economist using one source of data to predict the impact of a new health insurance policy might well develop projections which differ notably from projections made by an economist using another source of very similar information. The discrepancy among sources is marked in particular cases, and it is reasonable to use record linkage to bring some order out of this confusion.

To accommodate the problem, a massive Federal effort to reconcile conceptual differences among record contents has been mounted jointly by the U.S. Census Bureau, the Social Security Administration, and the Internal Revenue Service. The relevant data base includes the Bureau's 1973 Current Population Survey and administrative records from IRS and SSA files. The reconciliation has three immediate purposes: to understand the relationships among ostensibly identical categories of information maintained by each agency, to input resultant data into the SSA simulation models of the tax transfer system, and to assess relative biases in Census statistics. The reconciliation involves linking a strategied sample of records on individuals from the various sources, not linkage of the entire data bases. Preliminary results of the study reported by Herriott and

> Spiers (1975) suggest that census statistics on income are quite reliable for salaried employees and regular wage earners; the overlap between Census reports contents is about 96%. Income reports of the self-employed show somewhat less accuracy (90% agreement between Census and IRS); reports of interest and dividends made to census are considerably less reliable (less than 80% agreement) for most respondent groups.

As a result of such research, the models of economic systems employed by the U.S. Census Bureau and by the Social Security Administration (SSA) can be improved considerably when error rates based on IRS data can be recognized. The differential predictability of male and female incomes becomes more interpretable with evidence on differential accuracy in reporting such income to Census interviewers. The estimates of the impact of training on income become more reliable when corrected for base rate errors in reporting that income. And so on.

Similar benefits accrue from investigations of the differences in count data as a function of archival source.

> A study by Cobleigh and Alvey (1975), for example, shows that differences in legally defined coverage of the population by Census and by SSA produce a Census comparable to a universe which is about 94% of the SSA taxable earner's listings. Given a comparable universe, reports of average annual earnings from the two sources are in remarkable agreement except for very low and very high income groups. In the very low categories, SSA data show about 20% more wage earners than does the Census data; in the high income categories, however, the Census counts are 10-20% higher than Social Security reports. These latter differences are attributed by the authors to definitional differences and reporting irregularities including self-employment earnings not reportable to SSA, rounding error in self-reports to Census, late reporting to SSA, and to other factors.

Another type of enrichment involves the use of archival records for specialized research in which the record, though not disclosable by law or social custom to the social scientist, represents a key element in accomplishing applied research goals. Surrogates for the record may be sought, of course, but in the absence of any suitable substitute, it is often possible to capitalize effectively on restricted access records without according special privileges to the social scientist. For example, one of the peculiar and persistent tensions in our society involves the zealous efforts of the U.S. Internal Revenue Service to extract legitimate taxes from citizens and some citizens' equally strenuous efforts to avoid paying them. In an effort to clarify the conditions under which taxpayers will filfill their responsibility with somewhat less resistance (or at least dissatisfaction), Schwartz and Orleans (1976) mounted some experimental tests of those conditions to

compare relative rates of tax payments for a particular category of income.

> Taxpayers were assigned randomly to one of three advertising strategies, the strategies differing in respect to their emphasis in justifying payment of taxes. The first condition relied heavily on appeals to moral conscience, the second on threats of punitive legal action, and the third on threats of social embarrassment (tax evasion being a matter for public legal action). The objective of the experiments was to determine which types of appeal led to higher rates of the particular income. To do so credibly required that condition or form of appeal be linked with the individual's subsequent reports of income to the Internal Revenue Service. In order to link the two kinds of records (the researcher's record of condition and the IRS record of income) without breaching IRS rules on disclosure of records (which are confidential by law) and the researchers' rules concerning disclosure of their own records, a mutually insulated file approach, described in Section 5, was used. (The results of the experiment are interesting. Middle-income respondents react most to the threats of legal action; low-income groups respond most to appeals to moral conscience; the high-income groups were most affected by threats of social embarrassment).

The case for merging separate data sets into a permanent consolidated pool of data is based on the assumption that the pooled data will be a more informative basis for social research than separate files. Examples of these are few, however, because the difficulty of match-merging files, the differences in terminology, and differences in sample design and data collection procedures have inhibited many researchers from consolidating files. Moreover, it is difficult to anticipate the usefulness of linked files without actually trying the idea out on a small sample of records. Among the large-scale examples, the Wisconsin Assets and Income Studies Archive (Bauman, David, & Miller, 1970) illustrates what can be accomplished, however, Researchers appraise the effects of tax averaging proposals, changing incomes from retirements, capital gains income, and so on by simulating changes in tax laws, using the linked records as the raw material for analysis. Records from the Internal Revenue Service, Wisconsin tax records, the Social Security Administration, are combined in the file, without jeopardizing privacy of individuals on whom records are kept, to permit this research. The products of the research are predictions about the importance of changes in tax laws on individual income, strategy which attenuates the need to rely solely on anecdotal case study, intuition, and fragmented data as a basis for legislation in the tax area.

The more elaborate and more sensitive merged systems are found in the medical arena. Most involve both administrative and research information and, because they are recent systems, the benefits of pooling both kinds

of data are not yet clear. Nonetheless, good reviews of the early products of such work are available for social medicine, community health services systems, and the like (e.g. Acheson, 1967). Laska and Bank's (1975) description of the Rockland Institute's psychiatric information system is probably one of the best of its kind. There is a strong emphasis on legislative and technical safeguards for assuring the confidentiality of the records. There is a hard-nosed product orientation: aside from common demographic information, the system facilitates quality control over treatment, time series analyses, and projective studies of the incidence and development of mental illness, and permits some uncontrolled studies of the effectiveness of treatment. Perhaps most importantly, the system can be coupled neatly to experimental tests of alternative treatments to better understand whether and how well the treatments work (Endicott & Spitzer, 1975).

4. Privacy Implications: Private with Respect to Whom?

Any longitudinal research involves linking observations made on an individual (or some other unit of analysis) at one point in time with observations made at a second point. The average statistical relation derived from the constellation of individual observations is, as we've said, useful for description at least, and is often essential for planning and evaluating social programs, for understanding change in human behavior, and for building theory and simulation models. The linkage is usually but not always made on the basis of clear identification of the respondent. Insofar as the identified respondent does share information about himself, the sharing process may be regarded, in principle, as a depreciation of the individual's privacy. That depreciation may be quite innocuous in the sense that information disclosed is innocuous; or it may be controversial, as in longitudinal studies of mental health.

Similarly, correlational data analysis must often be based on linkage of records from different archives. And if that linkage is based on clear identification contained in each record, then privacy may be depreciated in principle here as well. The custodian of an administrative archive may, by permitting linkage, violate law at worst or social customs at best by disclosing records to a researcher for linkage, however worthwhile the purpose of linkage. There may be a similar breach of a promise of confidentiality for a researcher who discloses his own records on identifiable individuals to an administrative archive, for example, in order to verify his records against those maintained by the archive.

These implications are almost useless in developing general strategies for assuring individual privacy. For although disclosure of information may represent a depreciation of privacy in principle, the fact of the matter is that neither government, nor social or administrative science, nor the respondent could get on well without some exchange of information about individuals. Admitting this, the focus must change from absolute assurance of confidentiality to balancing social information against the privacy-related needs of the individual. One approach to

achieving that balance in a concrete way is to try to minimize depreciation of privacy without notably abridging our ability to collect meaningful data on human behavior. Doing so requires that we first identify the sources of risk in social research, then build mechanisms--procedural, statistical, and legal--to attenuate that risk.

We recognize, for example, that privacy may be reduced directly with respect to the social scientist. In the past, any such depreciation has been innocuous partly because social research itself has been fairly innocuous. But as applications of social research to social problems increase, as social scientists investigate more important or more controversial topics, the attention given to their inquiries are likely to increase. The import attached to relatively minor depreciation of privacy will increase. And so it becomes the social scientist's responsibility to develop mechanisms for minimizing the depreciation of privacy with respect to the researcher. I believe this in spite of the fact that no substantial risks to the respondent are usually engendered by survey research. The lack of risk is traceable to the researcher's lack of interest in making personal judgments about particular individuals and his interest in statistical analysis of the relevant data. Identifiers serve merely as an accounting device, rather than as a vehicle for administrative action against (or for that matter for) an individual. Nonetheless, if identifiers could somehow be eliminated in the research process, or if the tie between identifier and response could be made useless for making personal judgments about individuals, without damaging research objectives needlessly, then we would do so. Partial solutions to the problem of doing so (Section 5) have been developed partly as a matter of principle, and partly because risks of disclosure may be generated by persons or agencies other than the researcher.

It is clear, too, that fraudulent researcher, i.e., individuals posing as social scientists, can and occasionally do decieve citizens. They are motivated by financial gain (e.g., salesmen posing as pollsters), by pathological influences (e.g., rapists posing as survey interviewers, or at times, as policemen), or by other factors. In the interest of preserving the integrity of the profession and public trust in the social scientist, the social scientist must take some responsibility for protecting respondents against these infrequent but important dangers.

Social research records on identifiable individuals are often irrelevant for making <u>administrative</u> judgments about those individuals. We deal in samples rather than populations, and idiosyncratic ones at that. We deal with information which is usually <u>not</u> at the correct level of relevance or detail for administrative use. This partial relevance of research records on individuals usually serves as an inhibition against the appropriation of records for nonresearch purposes. Nonetheless, appropriation can and does occur. It may emerge under legal mandate as it has in the United States where, in a few instances, research records have been subpoenaed for use in judicial investigation of particular survey respondents. Exploitation may occur under legal traditions which are quite arbitrary and at times border on the capricious, as in some

Congressional investigating committee activity. Or, the exploitation may be quite illegal, as in the theft and use of research records for personal profit or for the purpose of embarrassing the respondent. The consequences to the respondent can be serious: social embarrassment, legal sanction, personal discomfort. The consequences for research are no less serious: its inhibition and abrogation, now and in the future.

These risks are in principle real, if in practice remote. And so they deserve attention too. In particular, it is reasonable to examine mechanisms which protect the respondent from capricious action by law enforcement agencies, from criminal action based on the information he provides to a researcher, and from other attempts to appropriate research records for nonresearch purposes. This is especially true for those cases in which the benefits of the research are likely to offset greatly the social benefits of legal appropriation of records.

5. Competing and Conjoint Approaches to Assuring Confidentiality of Response in Social Research[5]

The general implication of the preceding section is that we take as an objective reducing depreciation of privacy without severe abridgment of research goals. Accommodating this joint task is difficult but there have been a variety of efforts mounted recently to do so. The major strategic approaches can be grouped into three broad categories--procedural, statistical, and law-related--which we consider next. This examination is brief; details are given in Boruch (1976).

5.1 Procedural Approaches

For longitudinal data collected periodically within the same framework, the simple device of using alias identifiers is obvious is underutilized. The alias may be created by the respondent and used consistently in response to permit intrasystem linkage. It may be created by social scientists, provided to the respondent, then purged from the social scientists' files to achieve the same ends. To decentralize the process, some neutral brokerage agency (a census bureau, a nongovernmental agency) may similarly create an alias for the respondent and destroy its own records of any linkage between clear identification and alias.

The strategy has been field tested with some success in U.S. drug studies, political attitude surveys and the like. Aside from logistical problems, its major shortcomings are the limitations imposed on linking the data elicited under alias with any other existing data on individuals.

To accommodate some logistical problems as well as the limitation on intersystem linkage, procedures such as the link file system have been developed. In this technique, a dictionary of double aliases is created by the social scientist and given over for safekeeping to an independent agency. The decentralization of the process enhances physical security, and if the agency is legally entitled to resist governmental appropriation

of files, the procedure is legally secure. The dictionary is used as a basis for linking information which is periodically obtained from individuals. The main benefit of the strategy is that it reduces the social scientists' need to maintain longitudinal records on identified individuals, in general, and it reduces the time during which the social scientist has access to any given wave of data containing identifiers to to an arbitrarily short period (see Astin & Boruch, 1970).

For those cases in which records from different archives must be linked, a variety of methods have been developed to permit linkage without violating the customs or law governing linkage. Among the better known systems for doing so is the "mutually insulated file approach," used in the Schwartz-Orleans (1967) study cited earlier. Basically, the system involves two files of records operated under different auspices; all records are identified and there is some overlap between the samples of individuals on which the records are maintained. To accomplish the linkage, the first archive (assume it is the social scientist) cryptographically encodes the information portion of each record, producing a new file without meaning to any outsider, which is then transmitted to the record archive. The archive then matches the encoded records with its own records, based on the clear identifiers appearing in each record. Upon completion of the match, identifiers are deleted and the linked records are returned to the social scientist who then decodes relevant protions of the linked records and conducts his statistical analysis of the anonymous records. (See also Boruch, 1972, and Campbell et al., 1975).

These procedural approaches are simple, and in some cases, vulnerable to corruption. Nonetheless, they are useful in some, but not all research settings, to assure confidentiality of data with respect to the researcher and outsiders, and they can be tailored to accommodate longitudinal or correlational studies. Their refinement has been undertaken by both research community and the Federal bureaucracy to enhance the procedures' flexibility and protection level (Boruch, 1976). Some of the refinements depend on statistical approaches considered below.

5.2 Statistical Approaches

The devices just described are most often relevant to more impersonal forms of observation--questionnaires and the like--rather than to direct interview research. And in some instances, the logistical difficulties attached to their use are considerable. Partly for these reasons, it may be more appropriate to capitalize on one of the statistical strategies which have been developed to reduce depreciation in privacy. A variety of these approaches exists and these may be used alone or in conjunction with the procedural devices.

The best known class of approaches is the randomized response tactic currently under test and development by Greenberg in the United States, Dalenius (1975), Lanke, Swensson, Svensson, and Ericksson in Sweden, Warner in Canada, Moors in Holland, and others. In the simplest variation of the approach, the social scientist simultaneously presents

a sensitive inquiry to an individual, e.g., "Did you cheat on your income taxes this year?" and an insensitive one, e.g., "Do you prefer potatoes over noodles?" The individual is then instructed to roll a die and to respond to the first question if a one or two shows up, and to the second question if a three, four, five, or six shows. He is also told to refrain from giving the interviewer any indication of which question was answered. When the process is carried out on two large samples of individuals and the instructions are followed by the respondent, it is possible to estimate the proportion of individuals in the sample who have cheated on their income tax forms and the proportion who prefer noodles. In particular, given some simple laws of probability, the odds on answering one or the other question, the odds on answering one or the other question, and the observed proportion of Yes responses, the estimation is a matter of simple algebra.

The technique permits us to establish the statistical character of sensitive properties of groups of individuals. And moreover, it does so without disclosing to the social scientist any information about a particular individual. It has been field tested in drug studies, in fertility control studies and other areas, and those tests continue in the U.S., Canada, Sweden, and elsewhere. The basic method is being refined to make it more efficient in a statistical sense, more acceptable to the respondent in a social psychological sense, and less vulnerable to corruption in a legal sense.

A separate class of approaches is based on aggregation of response. The itdividual is asked not to respond individually to each of a set of questions but to respond in aggregated form to the set. In particular variations, for example, the respondent may add up numercial values corresponding to each answer of each question in a set. If "Yes" is assigned a value of 1 and "No" a value of -1, for example, the answer provided to a set of 10 questions each answerable with a Yes or No is a single number whose permissible range is -10 to +10. If numerical assignment is varied from one sample to the next, one needs only a little algebra--notably methods for solving a system of simultaneous equations--to estimate the proportions of individuals in the total sample who have each of the 10 properties.

Again, the technique permits one to elicit even sensitive information in direct interview situations without any deterministic linkage between an identified response to the researcher's question and the actual status of the individual. With some technical improvements, it probably can be applied to some longitudinal studies in which average relations among properties are essential.

The third and final class of statistical techniques which has received some attention is aggregation of the sample. The technique requires that one obtain data not on single identified individuals but rather on very small and carefully constructed clusters of individuals. If the cluster's composition remains the same over time, each cluster

can, under certain conditions, be regarded as a synthetic person, a composite of all the properties of the small set of individuals it comprises. Some informative data analyses can be conducted on those aggregates and, insofar as aggregation helps to assure anonymity of individual response, there is no depreciation of individual privacy.

The applications of sample microaggregation have so far been limited to economic research on commercial units. Banks, for example, may be reluctant to release information about their operations to any outside economist. They are willing, however, to have the social scientist analyze aggregates of banks in the interest of reconciling bank privacy with futures research. And indeed, a major system of data maintenance and dissemination has been built up on this theme by the University of Wisconsin (see Bauman, David, & Miller, 1970).

5.3 Approaches Based on Law and Government Practice

The final class of approaches to facilitating the privacy of the respondent in social research concerns formal legal action by legislators, the courts, or governmental executive agencies. Such action is taken to assure than when identifiable data must be collected for research purposes, the data will not be used for purposes other than research. As a practical matter, this means not only strengthening legal sanctions against criminal appropriation of research records, but also defining bounds on governmental appropriation of records. The actions are taken to reduce the likelihood that research records on identifiable individuals will be used to depreciate privacy any more than is normally required by research and to isolate that research against temporary threats, legal or otherwise, when the potential benefits of research justifies this course of action. The forms which such protection may take vary considerably, and so we describe only a few stereotypes here.

In some of the United States, public officials such as the governor are empowered by the state constitution or by legislative act to offer testimonial privilege to a social researcher. That privilege entitles the recipient to legally resist any legal effort to appropriate his records on identifiable individuals. The threat of appropriation may stem from a prosecutor's idea that he may use even an unwilling researcher as a criminal investigator. It may stem from arbitrary exercise of subpoena power by legislatures or the courts. In order to legally assure that data will not be so appropriated, and consequently to increase the likelihood that individuals will cooperate in the research, a governor may then provide testimonial privilege on an ad hoc basis. To take a specific example, the governor of Vermont gave such privilege to researchers and respondents who participated in roadside surveys of drivers. The survey objectives were to estimate the proportion of drinking drivers (bloodtests were given to drivers) and the privilege was essential in getting high cooperation rate. Drivers who were legally intoxicated were driven home by a policeman. No record of any identified individual's condition was lodged with any law enforcement agency or other government archive, though drivers would normally be prosecuted

under the law.

This sort of privilege can be applied in special cases where potential benefits of the survey are high and the relevant government executive is well enough informed to recognize the fact. However, we cannot always rely on expected benefits of research, for although some research may be important, it may also be risky with respect to its payoff. Nor can we always rely on the good offices of the public official, for the awarding of such privilege is discretionary and political factors may argue against it. In any event, discretionary privilege may be as susceptible to abuse from the naive researcher, just as it has been abused occasionally by some government executives.

Judicial discretion is another potential source of support for social scientists who, having collected identifiable data and having established a need for its maintenance, wish to secure it against non-research uses. In some cases, it has been possible for the scientist to legally resist a court-issued subpoena on grounds that the disclosure of identified records to thecourt would badly disable a major research effort. Evidence that breaches of confidentiality can be harmful to research efforts is readily available and can be used effectively to show cause why the records should not be used except in anonymous form. In fact, a similar line of argument has been used in a case involving the Negative Income Tax Experiments in New Jersey: The suspicion of fraud among people who happened to participate in the research led to a grand jury investigation and subpoena of research records on identified individuals.

Judicial discretion, like executive discretion, is by definition a bit arbitrary at best, and wildly unpredictable at worst. So its usefulness in protecting the confidentiality of data is not especially promising.

Legislative action in the form of concrete law is both feasible and, from the point of view of uniformity, very desirable. In particular, it is possible to build law to grant testimonial privilege to legitimate social scientists under well defined conditions and uniformly applied criteria. It is also possible to build into such law sanctions against the fraudulent researcher or the corrupt social scientist or the public official who might attempt to appropriate research data for research purposes.

The 1970 Drug Abuse Act and the 1970 Alcohol Abuse Acts, for example, each carry a statute which permits the Attorney General to accord privilege to social scientists who are funded by the government to conduct research on those topics. Under the Public Health Act, persons engaged in research on mental health, including the use of alcohol and other proactive crugs, can be accorded privilege by the Secretary of Health, Education, and Welfare to protect the privacy of individuals who are subjects of such research.

These are new laws, enacted specifically to assure the confidentiality of social research records on identifiable individuals. They represent a delimitation of power on governmental access to social research records, and a delimitation of the conditions under which the researcher may act. They represent a spirit of support for the social sciences as well as an appreciation for the negative impact which even legal appropriation of research records may exert on policy-relevant research. At least one such law has been tested by the courts, and it's intent has been reaffirmed in that arena as well.

References[6]

Acheson, E. D. Medical record linkage. London: Oxford University Press, 1967.

Astin, A. W., & Boruch, R. F. A "link system for assuring confidentiality of research data in longitudinal studies. American Educational Research Journal, 1970, 7(4), 615-624.

Barton, E. M., Plemons, J. K., Willis, S. L., & Baltes, P. B. Recent findings on adult and gerontological intelligence: Changing a stereotype of decline. American Behavioral Scientist, 1975, 19(2), 224-236.

Bauman, R. A., David, M. H., & Miller, R. F. Working with complex data files: The Wisconsin assets and income studies archive. In R. L. Biscoe (Ed.), Data bases, computers, and the social sciences. New York: Wiley-Interscience, 1970.

Bejar, I. I. Substudy 1a: Secondary analysis of the Cali tests of the nutrition and cultural enrichment program (Evaluation Research Memo #1IIB). Evanston, Illinois: Northwestern University, Department of Psychology, February 1976.

Birren, J. E. The psychology of aging. Englewood Cliffs, New Jersey: Prentice-Hall, 1962.

Boruch, R. F. Costs, benefits, and legal implications of methods for assuring confidentiality in social research (Evaluation Research Report NIE-020). Evanston, Illinois: Psychology Department, Northwestern University, 1974. Reprinted as: Statische und methodische proceduren zur Sicherung der Vertrau lichkeit bei Forshung. In A. Eser and K. F. Schumann (Eds.), Forshung im konflikt mit Recht und Ethik. Stuttgart: Ferdinand Enke Verlag, 1976.

Boruch, R. F. Strategies for eliciting and merging confidential social research data. Policy Sciences, 1972, 3(3), 275-297.

Boruch, R. F., & Creager, J. A. Measurement error in social and educational research (ACE Research Reports, Vol. 7, No. 2). Washington, D.C.: American Council on Education (One DuPont Circle), 1972.

Bureau of the Census. Some preliminary results from the 1973 CPS-IRS-SSA exact match study: Invited papers on the reconciliation of survey and administrative income distribution statistics through data linkage. Reproduced report. Washington, D.C.: U.S. Department of Commerce, Bureau of the Census, September 30, 1975.

Burstein, L. The Unit of analysis in educational research. Presented at the Annual Meetings of the American Educational Research Association,

1975, Washington, D.C.

Campbell, D. T. Administrative experimentation, institutional records, and nonreactive measures. In J. C. Stanley (Ed.), Improving experimental design and statistical analysis. Chicago: Rand McNally, 1967. Pp. 257-291.

Campbell, D. T. Assessing the impact of planned social change. In G. M. Lyons (Ed.), Social research and public policies: The Dartmouth-OECD Conference. Hanover, New Hampshire: University of New Hampshire Press, 1975. Pp. 3-45.

Campbell, D. T., & Boruch, R. F. Making the case for randomized assignment to treatments by considering the alternatives: Six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects. In A. Lumsdaine and C. A. Bennett (Eds.), Evaluation and measurement: Some critical issues in assessing social programs. New York: Academic Press, 1975. Pp. 195-296.

Campbell, D. T., Boruch, R. F., Schwartz, R. D., & Steinberg, J. Confidentiality preserving modes of access to files and to interfile exchange for useful statistical analysis. In A. Rivlin (Ed.), Report of the National Academy of Sciences, the Committee on Federal Agency Evaluation Policy: Protecting individual privacy in evaluation research. Washington, D.C.: NAS, 1975.

Campbeld, D. T., & Erlebacher, A. How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), Compensatory education: A national debate. New York: Brunner/Mazel, 1970.

Cobleigh, C., & Alvey, W. Validating reported social security numbers. In Proceedings of the American Statistical Association: Social Statistics Section. Washington, D.C.: ASA, 1975. P. 145-151.

Coronary Drug Project Research Group. The Coronary Drug Project: Design, Methods, and Baseline Results. American Heart Association Monograph, 1973, 38.

Dalenius, T. The invasion of privacy problem and statistics production. Forskningsprojektet, Fel I Undersokningar, Rapport Nr 61. Stockholm, Sweden: Statistika Institutionem, Stockholms Universitet, December, 1973.

Director, S. Evaluating the impact of manpower training programs. Ph.D. dissertation, Northwestern University (Evanston, Illinois), 1974.

Duncan, O. D. The 1970 Census: National uses--challenge and opportunity. In Proceedings of the American Statistical Association: Social Statistics Section, 1969. Washington, D.C.: ASA, 1966. Pp. 1-6.

Endicott, J., & Spitzer, R. L. Patient assessment and monitoring. In E. M. Laska and R. Bank (Eds.), Safeguarding psychiatric privacy. New York: John Wiley, 1975. Pp. 285-298.

Fagerlind, I. Formal education and adult earnings: A longitudinal study on the economic benefits of education. Stockholm: Almqvist and Wiksell, 1975.

Fisher, J. L. The uses of Internal Revenue Service data. In M. E. Borus (Ed.), Evaluating the impact of manpower programs. Lexington, Mass.: D. C. Heath, 1972. Pp. 177-80.

Gilmore, C. P. The real villain in heart disease. New York Times Magazine, March 25, 1973.

Goodman, L. Some alternatives to ecological correlation. American Journal of Sociology, 1959, 64, 610-625.

Goslin, D. A., & Bordier, N. Record keeping in elementary and secondary schools. In S. Wheeler (Ed.), On record: Files and dossiers in everyday life. New York: Russell Sage, 1969. Pp. 29-61.

Hansen, M. H., & Hargis, B. J. Census Bureau uses of tax data. In Proceedings of the Business and Economic Statistics Section: American Statistical Association. Washington, D.C.: ASA, 1966. Pp. 160-164.

Heber, R., Garber, H., Harrington, S., Hoffman, C., & Falender, C. Rehabilitation of families at risk for mental retardation. Madison, Wisconsin: University of Wisconsin, Rehabilitation Research and Training Center, 1972.

Heller, R. N. The uses of social security administration data. In M. E. Borus (Ed.), Evaluating the impact of manpower programs. Lexington: Mass.: D. C. Heath, 1972. Pp. 197-201.

Herriot, R. A., & Spiers, E. F. Measuring the impact on income statistics of reporting differences between the Current Population Survey and administrative sources. Proceedings of the Social Statistics Section, American Statistical Association. Washington, D.C.: ASA, 1975.

Horvitz, D. G. Problems in designing interview surveys to measure population growth. Proceedings of the Social Statistics Section, American Statistical Association. Washington, D.C.: ASA, 1966. Pp. 245-249.

Jabine, T. B., & Rothwell, N. D. Split-panel tests of census and survey questionnaires. 1970 Proceedings of the Social Statistics Section, American Statistical Association. Washington, D.C.: 1970. Pp. 4-13.

James, K. E. Regression toward the mean in uncontrolled clinical studies. Biometrics, 1973, 29, 121-130.

Janson, C. G. Project Metropolitan: A longitudinal study of a Stockholm cohort (Research Report No. 1). Stockholm, Sweden: Department of Sociology, Stockholm University, 1975.

Kannel, W. B., Dawber, T. R., Kagan, A., Revotskie, N., & Stokes, J. Factors of risk in the development of coronary heart disease--Six-year-followup experience: The Framingham Study. Annals of Internal Medicine, 1961, 55(33).

King, L. S. (Ed). A history of medicine. Middlesex: Penguin, 1971.

Klevmarken, A. Statistical methods for the analysis of earnings data. Stockholm: Alqvist and Wiksell, 1972.

Laska, E. M., & Bank, R. (Eds.). Safeguarding psychiatric privacy. New York: John Wiley, 1975.

Magidson, J., Barnow, B. S., & Campbell, D. T. Correcting the under-adjustment bias in the original Headstart evaluation (Evaluation Research Report No. 2JM). Evanston, Illinois: Psychology Department, Northwestern University, 1976.

Magnusson, D., Duner, A., & Zetterblom, G. Adjustment: A longitudinal study. Stockholm, Sweden: Almqvist and Wiksell, 1975.

Marks, E. S., & Waksberg, J. Evaluation of coverage in the 1960 Census of the Population through case by case checking. Proceedings of the American Statistical Association: Social Section. Washington, D.C.: ASA, 1966. Pp. 62-70.

Mednick, S. A., & McNeil, T. F. Current methodology in research on the etiology of schizophrenia. Psychological Bulletin, 1968, 70, 681-693.

Mednick, S. A., Schulsinger, F., & Garfinkle, R. Children at risk for schizophrenia: Predisposing factors and intervention. In M. L. Keitzman, S. Sutton, and J. Zubin (Eds.), Experimental approaches to psychopathology. New York: Academic Press, 1975. Pp. 451-464.

Miller, H. P., & Hornseth, R. A. Cross-sectional versus cohort estimates of lifetime income. Proceedings of the American Statistical Association: Social Statistics Section. Washington, D.C.: ASA, 1970. Pp. 339-341.

Milles, M. T., & Kilss, B. Exploration of differences between linked Social Security and Internal Revenue Service wage data for 1974. Proceedings of the Social Statistics Section, American Statistical Association. Washington, D.C.: American Statistical Association,

1975.

Nesselroade, J. R., & Baltes, P. D. Adolescent personality development and historical change: 1970-1972. Monographs of the Society for Research in Child Development, 1974, 39(1).

Parnes, H. S. The National Longitudinal Surveys: New vistas for labor market research. Journal of the American Economic Association, 1975, 65(2) 244-249.

Potter, R. G., Chow, L. P., Jain, A. K., & Lee, C. H. Social and demographic correlates of IUCD effectiveness: The Taichung IUCD medical follow-up study. Proceedings of the American Statistical Association: Social Statistics Section. Washington, D.C.: ASA, 1966. Pp. 272-277.

Ramcharan, S., Cutler, J. L., Feldman, R., Siegelaub, A. B., Campbell, B., Friedman, G. D., Dales, L. G., & Collen, M. F. Multiphasic checkup evaluation study: 2. Disability and chronic disease after seven years of multiphasic health checkups. Preventive Medicine, 1973, 2, 207-220.

Reeder, L. G., et al. Advances in health survey research: Proceedings of a national invitational conference. Rockville, Maryland: National Center for Health Statistics, 1975.

Robertson, L. S., Kelley, A. B., O'Neill, B., Wixom, C. W., Eiswirth, R. S., & Haddon, W. A controlled study of the effect of television messages on safety belt use. Washington, D.C.: Insurance Institute for Highway Safety (Watergate Six Hundred), 1972.

Robinson, W. S. Ecological correlations and behavior of individuals. American Sociological Review, 1950, 15, 351-357.

Schaie, K. W. A general model for the study of developmental problems. Psychological Bulletin, 1965, 64(2), 92-107.

Scheuren, F., & Alvey, W. Selected bibliography on matching. In F. Scheuren et al., Exact match research using the 1973 Current Population Survey. Studies from Interagency Data Linkages (Report No. 4). Washington, D.C.: Office of Research and Statistics, Social Security Administration, 1975.

Scheuren, F., Herriot, R., Vogel, L., Vaughan, D., Kilss, B., Tyler, B., Cobleigh, C., & Alvey, W. Exact match research using the March 1973 Current Population Survey--Initial states. Studies from Interagency Matching (Report No. 4). Washington, D.C.: Office of Research and Statistics, Social Security Administration, 1975.

Schwartz, R. D., & Orleans, S. On legal sanctions. University of Chicago Law Review, 1967, 34, 282-300.

Sinesterra, L., McKay, H., & McKay, A. Stimulation of intellectual and social competence in Colombian preschool children affected by multiple deprivations of depressed urban environments (Progress Report, mimeo). Cali, Colombia: University Center for Child Development, Human Ecology Research Station, Universidad del Valle, 1971.

Steinberg, J., & Pritzker, L. Some experiences with and reflections on data linkage in the United States. Bulletin of the International Statistical Institute, 1969, 42 (Book 2), 786-805.

Wall, W. D., & Williams, H. L. Longitudinal studies and the social sciences. London: Heineman, 1970.

Wohlwill, J. F. Methodology and research strategy in the study of developmental change (ETS Research Memorandum RM-69-24). Princeton: Educational Testing Service, December 1969.

Wohlwill, J. F. The age variable in psychological research. Psychological Bulletin, 1970, 77(1), 49-64.

Young, A. F., Selove, J. M., & Koons, D. A. Measuring quality of housing. Proceedings of the American Statistical Association: Social Statistics Section, 1966. Washington, D.C.: ASA, 1966. Pp. 33-42.

Footnotes

1. Background research for this paper was supported under a contract (NIE-C-74-0115) with the National Institute of Education.

2. For empirical data, theory, and policy implications of longitudinal studies in development, see Schaie (1965), Wohlwill (1969, 1970), and Magnusson, Duner, and Zetterblom (1975). The results of a variety such studies in Britain, the United States, France, and Germany are summarized by Wall and Williams (1970).

3. Cohort effects have been recognized only recently by commercial market researchers as an important variable in predicting and explaining the demand for certain consumer goods. Systematic cohort variation in what is regarded as a luxury item, for example, has some important implications for planning the allocation of an industry's manufacturing resources (see Business Week, January 12, 1976, pp. 74-78).

4. Conducting special social surveys to assess the quality of routinely issued governmental statistics is not a new idea. Neither is government's attempt to suppress the results of special surveys novel. See Boruch (1976) for a review of suppression efforts at the local, regional, and national level.

5. For a detailed examination of the benefits, shortcomings, vulnerability, and legal implications of some of these strategies, see Boruch (1974), and Campbell, Boruch, Schwartz, and Steinberg (1975).

6. This list contains some items which have not been specifically cited in the text.